

## DOCUMENT RESUME

ED 366 653

TM 021 077

AUTHOR Sheehan, Kathleen M.; And Others  
TITLE A Diagnostic Classification Model for Document Processing Skills.  
INSTITUTION Educational Testing Service, Princeton, N.J.  
SPONS AGENCY Office of Naval Research, Arlington, VA. Cognitive and Neural Sciences Div.  
REPORT NO RR-93-39-ONR  
PUB DATE Oct 93  
CONTRACT N00014-90-J-1307  
NOTE 43p.  
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.  
DESCRIPTORS \*Classification; \*Diagnostic Tests; Educational Diagnosis; Knowledge Level; Mastery Learning; Models; National Surveys; Performance; \*Problem Solving; Responses; Skills; Tables (Data); Test Items  
IDENTIFIERS Document Literacy Skills; \*Missing Data; National Assessment of Educational Progress; \*Rule Space Model

## ABSTRACT

This paper introduces a modification to the Rule Space diagnostic classification procedure to allow for processing of response vectors that contain missing data. Rule Space is an approach to diagnostic classification that involves characterizing examinees' performances in terms of an underlying cognitive model of generalized problem-solving skills. It has two components: (1) a procedure for determining a comprehensive set of knowledge states, where each state is characterized in terms of a unique subset of mastered skills; and (2) a procedure for classifying examinees into one or another of the specified states. Missing data are expected to be a common problem for this approach because, although the procedure for determining the comprehensive set of knowledge states requires a large pool of items, the procedure for examinee classification can be performed with smaller (less expensive) item subsets. This approach to diagnostic classification is illustrated with data collection in the Survey of Young Adult Literacy, a nationwide survey of literacy skills conducted by the National Assessment of Educational Progress (NAEP). The study has four figures and seven tables. An appendix contains a proof of an aspect of the model. (Contains 18 references.) (Author/SLD)

\*\*\*\*\*  
\* Reproductions supplied by EDRS are the best that can be made \*  
\* from the original document. \*  
\*\*\*\*\*

TM

ED 366 653

U.S. DEPARTMENT OF EDUCATION  
Office of Educational Research and Improvement  
EDUCATIONAL RESOURCES INFORMATION  
CENTER (ERIC)

RR-93-39-ONR

- ☒ This document has been reproduced as received from the person or organization originating it
- ☐ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

## A DIAGNOSTIC CLASSIFICATION MODEL FOR DOCUMENT PROCESSING SKILLS

Kathleen M. Sheehan  
Kikumi K. Tatsuoka  
Charles Lewis

This research was sponsored in part by the  
Cognitive Science Program  
Cognitive and Neural Sciences Division  
Office of Naval Research, under  
Contract No. N00014-90-J-1307  
R&T 4421559

Kathleen M. Sheehan, Principal Investigator



Educational Testing Service

Princeton, NJ

October 1993

Reproduction in whole or in part is permitted  
for any purpose of the United States Government.

Approved for public release; distribution unlimited.

BEST COPY AVAILABLE

10021877  
ERIC  
Full Text Provided by ERIC

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
<small>This report on burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302 and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.</small>				
1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE 8/30/93	3. REPORT TYPE AND DATES COVERED Final, April 1989 - August 1993		
4. TITLE AND SUBTITLE A Diagnostic Classification Model for Document Processing Skills		5. FUNDING NUMBERS C-N00014-90-J-1307 61153 N RR 04204-01 R & T 4421559		
6. AUTHOR(S) Kathleen Sheehan, Kikumi Tatsuoka, Charles Lewis				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Educational Testing Service Rosedale Road Princeton, NJ 08541		8. PERFORMING ORGANIZATION REPORT NUMBER ETS RR-93-39-ONR		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Cognitive Sciences (1142CS) Office of Naval Research 800 N. Quincy Street Arlington VA 22217-5000		10. SPONSORING/MONITORING AGENCY REPORT NUMBER		
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release Distribution unlimited			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) <p>This paper introduces a modification to the Rule Space diagnostic classification procedure which allows for processing of response vectors containing missing data. Rule Space is an approach to diagnostic classification which involves characterizing examinees' performances in terms of an underlying cognitive model of generalized problem-solving skills. It has two components: (1) a procedure for determining a comprehensive set of knowledge states, where each state is characterized in terms of a unique subset of mastered skills; and (2) a procedure for classifying examinees into one or another of the specified states. Missing data is expected to be a common problem for this approach because, although the procedure for determining the comprehensive set of knowledge states requires a large pool of items, the procedure for examinee classification can be performed with smaller (less expensive) item subsets. This approach to diagnostic classification is illustrated with data collection in the Survey of Young Adult Literacy, a nationwide survey of literacy skills conducted by the National Assessment of Educational Progress (NAEP)</p>				
14. SUBJECT TERMS Cognitive diagnosis, classification, IRT			15. NUMBER OF PAGES 32	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL	

# **A Diagnostic Classification Model for Document Processing Skills**

May 1993

Kathleen Sheehan,  
Kikumi Tatsuoka and Charles Lewis

Educational Testing Service  
Princeton, NJ 08541

Copyright © 1993. Educational Testing Service. All rights reserved.

## **A Diagnostic Classification Model For Document Processing Skills**

### **Abstract**

This paper introduces a modification to the Rule Space diagnostic classification procedure which allows for processing of response vectors containing missing data. Rule Space is an approach to diagnostic classification which involves characterizing examinees' performances in terms of an underlying cognitive model of generalized problem-solving skills. It has two components: (1) a procedure for determining a comprehensive set of knowledge states, where each state is characterized in terms of a unique subset of mastered skills; and (2) a procedure for classifying examinees into one or another of the specified states. The procedure for determining a comprehensive set of knowledge states is based on the Boolean descriptive function given in Tatsuoka (1991). The procedure for classifying examinees involves comparing examinees' scored response vectors to the patterns expected within each of the specified knowledge states (Tatsuoka, 1983, 1985, and 1987). Missing data is expected to be a common problem for this approach because, although the procedure for determining the comprehensive set of knowledge states requires a large pool of items, the procedure for examinee classification can be performed with smaller (less expensive) item subsets. This approach to diagnostic classification is illustrated with data collected in the Survey of Young Adult Literacy, a nationwide survey of literacy skills conducted by the National Assessment of Educational Progress (NAEP) in 1985.

## **A Diagnostic Classification Model For Document Processing Skills**

Many procedures for diagnostic classification require specification of the universe of procedural bugs accounting for examinees' errors. Diagnostic classification is subsequently performed by comparing an examinee's observed performance on a representative set of items to the performances expected under each of the specified buggy procedures. When a good match is found, the examinee is classified as having that particular bug.

For problems of typical size and complexity, however, the bug enumeration approach may not be feasible. An alternative, less fine-grained approach to diagnostic classification involves characterizing examinees' performances in terms of an underlying cognitive model of generalized problem-solving skills. Examinees' observed performances can then be compared to the performances expected at different mastery levels defined with respect to the underlying skills. Thus, the problem of enumerating all possible buggy procedures is replaced by two new problems: (1) identifying the unobservable, cognitive skills underlying performance, and (2) translating these skills into a comprehensive set of diagnostically relevant knowledge states. These two new problems may be more amenable to solution, especially in situations where a cognitive theory of performance is already available.

In this paper we assume that the cognitive skills underlying performance have already been identified and describe (1) a procedure for determining a comprehensive set of diagnostically relevant knowledge states; and (2) a procedure for classifying examinees' observed response vectors into one or another of the specified knowledge states. The procedure for determining a comprehensive set of knowledge states is based on the Boolean descriptive function given in Tatsuoaka (1991). The examinee classification procedure is a modification of the Rule Space classification procedure which allows for processing of response vectors containing missing data. Missing data is expected to be a common problem for these procedures because the method for determining a comprehensive set of knowledge states is defined with respect to a specific item pool. As will be seen later, this encourages the use of large diverse item pools for knowledge state definition and smaller (less expensive) item subsets for examinee classification.

This new approach to diagnostic classification is described in the following sections. The procedure for determining a comprehensive set of knowledge states is presented first. Second, the Rule Space classification procedure is described. Third, differences between this approach and an approach based on latent class analysis are presented. Fourth, modifications to the Rule Space classification procedure which were developed to handle the expected missing data problem are described. Fifth, this approach is applied to the problem of diagnosing document processing skills. The data available for the application were collected in the Survey of Young Adult Literacy, a nationwide survey of literacy skills conducted by the National Assessment of Educational Progress (NAEP) in 1985. The unobservable ordinally-scaled variables assumed to be underlying performance on document processing tasks were derived from the work of Kirsch and Mosenthal (1990) who identified features of

the items which were found to be highly correlated with proficiency in the domain. Finally, two new methods for analyzing the classification results are presented.

### **Determining a Comprehensive Set of Knowledge States**

The process of determining a comprehensive set of knowledge states in a domain of interest begins with the specification of the elementary cognitive skills needed for mastery of the domain. In Birenbaum, Kelly and Tatsuoka (1992), for example, proficiency in the domain of elementary algebra is broken down into a set of 11 component skills including: (1) ability to apply the distributive law; (2) ability to apply arithmetic order of operations laws; (4) ability to recognize when it makes sense to subtract a term from both sides of an equation; and (5) ability to recognize when it makes sense to divide both sides of an equation by the coefficient of  $x$ . (For a list of the remaining seven skills, see Birenbaum et. al., 1992.) Thus, although proficiency in solving elementary algebra problems is generally thought of as a unidimensional trait, a significant proportion of the variation in that trait may be accounted for by a diverse set of more elementary skills.

Note that the elementary algebraic skills listed above are all reported in a dichotomized fashion. Also, they are all diagnostically relevant in the sense that knowledge of the subset of skills possessed by an examinee constitutes information which one would expect to find useful for remediation. These two characteristics of skills (i.e. ability to dichotomize and relevance to remediation) are required for successful application of the diagnostic classification procedures described below.

Once the elementary cognitive skills underlying proficiency in the domain of interest have been identified, a comprehensive set of latent cognitive states can be determined by listing all possible subsets of skills mastered. For example, consider a model consisting of two skills  $A_1$  and  $A_2$ . The set of all possible subsets of these skills consists of the following four elements:

- 1.) The examinee has mastered both  $A_1$  and  $A_2$ .
- 2.) The examinee has mastered  $A_1$  but has not mastered  $A_2$ .
- 3.) The examinee has mastered  $A_2$  but has not mastered  $A_1$ .
- 4.) The examinee has not mastered  $A_1$  or  $A_2$ .

Thus, the universe of all possible latent cognitive states can be specified in terms of a set of four states. Due to the combinatorial nature of this problem, however, this method of determining the universe of latent cognitive states will not always be feasible. In the document processing illustration presented below, for example, the cognitive model yielded a total of 22 skills. The corresponding set of all possible subsets of skills mastered would include  $2^{22} = 4.2 \times 10^6$  elements, too many to consider, much less enumerate.



An alternative procedure for specifying the universe of all possible latent cognitive states is described in Tatsuoka (1991). (In Tatsuoka, the elementary cognitive skills are termed attributes. In this paper, the terms attribute and elementary cognitive skill are used interchangeably.) In this alternative procedure, characteristics of the available item pool are exploited to select a subset of states for further consideration. This is accomplished in two steps. First, in a step inspired by the work of Scheiblechner (1972) and Fischer (1973), each item in the pool is classified as to the subset of skills required for successful completion. This classification must be performed by someone who is familiar both with the items and with the cognitive model proposed for solving the items. The result is an incidence matrix  $Q$  whose order is the number of attributes ( $K$ ) by the number of items ( $n$ ). If item  $j$  requires mastery of skill  $k$  then  $Q_{kj}=1$ , otherwise  $Q_{kj}=0$ . Second, a Boolean descriptive function (BDF) is used to extract only those combinations of attributes which are represented in the available item pool. For example, consider a model involving ten attributes,  $A_1$  through  $A_{10}$ . If every item that required mastery of  $A_{10}$  also required mastery of  $A_9$ , then all states combining mastery of  $A_{10}$  with nonmastery of  $A_9$  would be excluded from the set of selected states (regardless of the mastery status specified for the remaining eight attributes).

As this example shows, states that are psychologically and logically valid but not distinguishable from the available item pool would not be extracted by the BDF. Thus, this procedure encourages the use of a large diverse item pool. For best results, the pool should contain at least one item tapping each expected combination of skills. Note that the BDF only requires that the items be classified according to required attributes. Thus, a comprehensive set of knowledge states can be determined without actually administering all of the items in the pool.

### Classifying Observed Response Patterns

The classification procedure described here involves comparing examinees' scored response patterns, ( $X_i=[x_{i1},...,x_{in}]$ , where  $x_{ij}$  is the response of the  $i$ th examinee to the  $j$ th item, 1 if correct, 0 if incorrect, and  $n$  is the number of items in the entire item pool) to the patterns expected within each of the specified knowledge states. First, each state is characterized by an ideal item response vector indicating the subset of items that would be successfully solved by an examinee in that state ( $X_s=[x_{s1},...,x_{sn}]$ ,  $s=1,...,S$ ). The process of associating an ideal item response vector with a particular state is fairly straightforward: when the incidence matrix indicates that a particular item requires a particular combination of attributes, the ideal response to that item will be correct for all states having that combination of attributes and incorrect for all others. Once an ideal response pattern has been defined for each state, the Rule Space classification procedure (Tatsuoka, 1985, 1987) can be used to classify examinees' observed response patterns as indicating the pattern of attribute mastery associated with one or another of the specified cognitive states.

A unique feature of the Rule Space classification procedure is that the comparison of examinees' observed response patterns to the various ideal response patterns is performed in a

reduced space that has only two dimensions. These two dimensions were selected to capture variation in the response patterns that would be considered important from the vantage point of Item Response Theory (IRT). The first dimension corresponds to the IRT proficiency estimate  $\hat{\theta}$ . (Hereafter,  $\hat{\theta}$  will be written as  $\theta$  for simplicity.) This dimension is important because it describes variation in the response patterns that can be attributed to differences in examinee proficiency levels. The second dimension corresponds to the variable  $\zeta$  which is an index of how unusual a particular item response pattern is (Tatsuoka, 1984, 1985). The  $\zeta_i$  associated with a particular response vector  $X_i$  is calculated as follows

$$\zeta_i = \frac{f(\theta_i, X_i)}{\sqrt{\text{Var } f(\theta_i, X_i)}}$$

$$\text{where } f(\theta_i, X_i) = \sum_{j=1}^n (P_j(\theta_i) - x_{ij}) (P_j(\theta_i) - T(\theta_i)) ,$$

$$\text{Var } f(\theta_i, X_i) = \sum_{j=1}^n P_j(\theta_i) (1 - P_j(\theta_i)) (P_j(\theta_i) - T(\theta_i))^2$$

$$\text{and } T(\theta_i) = \frac{1}{n} \sum_{j=1}^n P_j(\theta_i) .$$

In the above equations,  $P_j(\theta_i)$  is the probability of a correct response to the  $j^{\text{th}}$  item by the  $i^{\text{th}}$  examinee (as determined from the assumed IRT model), and  $T(\theta_i)$  is the average probability of a correct response, calculated over all items. Note that  $P(\theta_i) - X_i$  measures the deviation of the item response vector  $X_i$  from its expected value  $P(\theta_i)$ , and  $P(\theta_i) - T(\theta_i)$  measures the deviation of the expected value of the response vector  $X_i$  from the overall average probability of a correct response at  $\theta_i$ .

To illustrate the importance of  $\zeta$  in comparing different item response patterns, Table 1 lists sample  $\zeta$  values for a five-item test calibrated under the Rasch model with difficulty parameters of -2, -1, 0, 1 and 2. Each of the patterns listed in the table corresponds to a number correct score of 3, and thus, has an associated IRT proficiency estimate of  $\theta=.51$ . The table shows two things: first, the  $\zeta$  variable has been successful at capturing variation in the response patterns which was not captured by the proficiency estimate  $\theta$ ; and second, the  $\zeta$  values can be used to order the response patterns from those conforming to a Guttman pattern ( $\zeta=-.85$ ) to those conforming to a reverse Guttman pattern ( $\zeta=6.10$ ). Thus, another way to think about  $\zeta$  is that it indicates how well respondents' patterns accord with the assumed IRT model; low values indicate good fit (signaled by a Guttman pattern) and high values indicate poor fit (signaled by a reverse Guttman pattern).

Tatsuoka (1983) has noted that "similar" response patterns will have similar values of  $\theta$  and  $\zeta$ . Thus, one can evaluate the "similarity" of response patterns by mapping them into the two dimensional space formed by the Cartesian product of  $\theta$  and  $\zeta$ . This space is termed

the rule space. We note here that the mapping from response pattern to  $\zeta$  will only be one-to-one under certain conditions. (Dibello & Baillie, 1991). However, a one-to-one mapping can be assumed for most Rule Space applications because the conditions under which the mapping will not be one-to-one, as derived in Dibello & Baillie, will rarely be found among data which fit an IRT model.

---

---

Insert Table 1 Here

---

---

After the ideal item response vectors associated with each of the possible latent cognitive states have been mapped onto the two-dimensional rule space, determination of skill mastery for a particular examinee can proceed according to the following steps. First, the examinee's observed item response vector is also projected onto the two-dimensional rule space. Second, a subset of admissible states is determined by applying an admissibility criterion to each possible state. The admissibility criterion is defined in terms of the Mahalanobis distance ( $D_{is}$ ) between the examinee's point in the rule space ( $X_i$ ,  $i=1,\dots,N$ ) and the points associated with each of the ideal item response vectors ( $X_s$ ,  $s=1,\dots,S$ ). In particular, State  $s$  is admissible if

$$D_{is}^2 < \chi_{(2,\alpha)}^2$$

$$\text{where } D_{is}^2 = (\theta_i - \theta_s)^2 I(\theta_s) + (\zeta_i - \zeta_s)^2,$$

and  $I(\theta_s)$  is the Fisher information associated with the estimate  $\theta_s$  and  $\chi_{(2,\alpha)}^2$  is the  $\alpha$ -quantile of a chi-square random variable with 2 degrees of freedom. (We also say that State  $s$  is contained in the examinee's admissibility region.) Thus, an examinee's admissibility region contains the subset of states whose ideal item response vectors most closely resemble the examinee's observed item response vector, as determined by the Mahalanobis distance criterion.

Let  $r$  be a state in the admissibility region determined for examinee  $i$ . The posterior probability that this examinee has the pattern of skill mastery associated with State  $r$  can be determined as follows

$$P(r|\theta_i, \zeta_i) = \frac{P(\theta_i, \zeta_i|\theta_r, \zeta_r) P(r)}{\sum_{s=1}^S P(\theta_i, \zeta_i|\theta_s, \zeta_s) P(s)}$$

where  $P(r)$  and  $P(s)$  represent prior probabilities for states  $r$  and  $s$  ( $s=1, \dots, S$ ) respectively. The conditional probability,  $P(\theta_i, \zeta_i | \theta_r, \zeta_r)$  is taken to be bivariate normal with mean

$$M = \begin{bmatrix} \theta_r \\ \zeta_r \end{bmatrix}$$

and variance-covariance matrix

$$\Sigma = \begin{bmatrix} \frac{1}{I(\theta_r)} & 0 \\ 0 & 1 \end{bmatrix}$$

At this point, two alternative methods for determining attribute mastery classifications are available. First, in a manner similar to a latent class analysis, one could select the best available description of the examinee's true mastery profile by selecting that state with the highest posterior probability. For example, if State  $r$  had the highest posterior probability of all the states in the examinee's admissibility region, then the examinee would be classified into State  $r$ , or in other words, he or she would be diagnosed as having the pattern of attribute mastery associated with State  $r$ . Alternatively, it may be more appropriate to estimate an attribute mastery vector for each examinee by taking a weighted average of the attribute mastery designations associated with each of the states in the admissibility region. As an example, consider an admissibility region consisting of two states with the following attribute mastery patterns: {State  $r$ : 100} and {State  $q$ : 110}. A weighted average of these mastery designations would provide the following vector of attribute mastery values:

$$\begin{aligned} P(A_1) &= 1.0 \\ P(A_2) &= P(q|\theta_i, \zeta_i) / [P(r|\theta_i, \zeta_i) + P(q|\theta_i, \zeta_i)] \\ P(A_3) &= 0.0 \end{aligned}$$

where  $P(r|\theta_i, \zeta_i)$  and  $P(q|\theta_i, \zeta_i)$  represent posterior probabilities for States  $r$  and  $q$ , respectively. Note that, in this alternative method, an examinee's mastery status is described probabilistically rather than absolutely. This alternative method may be more or less appropriate depending on the ways in which the classification results are to be used.

### Comparison to Latent Class Analysis

Since latent class analysis also has as its objective the classification of observed response vectors into one or more of a set of latent cognitive states where each state is characterized by an idealized pattern of correct and incorrect responses (Lazarsfeld and Henry,

1960; Goodman, 1974; Macready and Dayton, 1980) it is useful to examine the differences between these two approaches.

A unique feature of the latent class approach is that each latent cognitive state is additionally characterized by a set of conditional probabilities,  $\alpha_s$  and  $\beta_s$ . The probability  $\alpha_s$  is the conditional probability of a correct response to any item for which the idealized pattern  $X_s$  indicates a correct response, given that the examinee has the pattern of skill mastery associated with State  $s$ . Similarly,  $\beta_s$  is the conditional probability of a correct response to any item for which the idealized pattern  $X_s$  indicates an incorrect response, given that the examinee has the pattern of skill mastery associated with State  $s$ . From specified values of  $\alpha_s$  and  $\beta_s$ , it is possible to calculate  $p_s(X_i)$ , the posterior probability that an examinee belongs to latent class  $s$ , i.e. has the pattern of skill mastery associated with State  $s$ , given their observed pattern of correct and incorrect responses,  $X_i$ . Diagnostic classification can then be performed by classifying each examinee into the class with the highest posterior probability. Note that the Rule Space approach does not require the specification of conditional probabilities  $\alpha_s$  of  $\beta_s$ .

A second difference between the Rule Space approach and a latent class approach is that the latent class approach provides very little guidance in the specification of knowledge states. By contrast, in the Rule Space approach, the comprehensive set of knowledge states is completely determined by the specification of the underlying cognitive model and the characteristics of the available item pool. If the item pool is developed to contain items tapping each of the relevant cognitive skills, then all of the relevant knowledge states will be extracted.

A third way in which the current approach differs from a latent class approach is that the current approach provides detailed information about which skills the examinee has and has not mastered. By contrast, the latent class approach merely provides information about which state the examinee has been classified into. Since states are not necessarily broken down into their more elementary cognitive components, the link to an effective remediation strategy is not as direct.

### The Missing Data Modification

In the classification procedure outlined above, each examinee's observed item response vector is compared to a single set of ideal item response vectors. Thus, it is assumed that each examinee is presented the same subset of items. In some testing situations, however, it will not be possible to administer the entire item pool to each examinee. In many large-scale testing programs, for example, multiple-matrix item sampling designs are used to efficiently measure population characteristics from sparse matrix samples of item responses. (Mislevy, Beaton, Kaplan and Sheehan, 1992). In these designs, different subsets of items are presented to different subsets of examinees. The NAEP data analyzed below provides an example. These data were collected under an item sampling design, called balanced incomplete block



(BIB) spiralling, in which the item pool is first divided into blocks and subsets of blocks are then grouped into test booklets such that each pair of blocks appears together in exactly one booklet. This design violates the assumption of no missing data since each examinee is only presented a subset of the entire item pool. This section describes a modification to the rule space procedure which was developed to allow processing of data sets containing missing item responses. Note that the procedure for determining a comprehensive set of knowledge states is not affected by this modification, since that procedure requires only the incidence matrix, not examinee response vectors.

To allow for different patterns of missing responses among different examinees, the missing data modification described here has been tailored to match the particular set of items presented to an examinee. That is, only those items which were actually administered to an examinee are considered during the classification of that particular examinee. This is accomplished in two steps: first, all 'not presented' items are masked out of the examinee's observed item response vector; and second, these same items are masked out of each state's ideal item response vector. Classification decisions are then made by comparing the examinee's reduced item response vector to each of the states' reduced ideal item response vectors. That is, both the examinee's reduced item response vector and each of the reduced ideal item response vectors are projected into the two-dimensional rule space and the Bayes decision rule described previously is applied. Note that this modification involves a great deal of additional computation since the ideal item response vectors associated with each state must be projected into the rule space  $N$  times, once for each examinee. By contrast, in the original rule space procedure the ideal item response vectors are projected into the rulespace once and this single projection is assumed to serve for all examinees.

Note that this approach does not involve any assumptions about the examinee's probable responses to missing items. Rather, a masking procedure is used to remove not-presented items from consideration entirely. An unintended result of the masking of ideal item response vectors is that two or more states may then be projected onto identical points in the rule space. When this occurs, it is an indication that the sampling design had not allowed for testing of all relevant attributes. To illustrate this point, consider a five-item test in which each item tests mastery of a single attribute. Two possible ideal item response vectors for this test are listed below.

- o Ideal response pattern for State r: 10100
- o Ideal response pattern for State q: 10101 .

Since States r and q differ only in their response to item 5, the reduced ideal item response vectors associated with these two states will be indistinguishable with respect to any item subset which does not include item 5. Thus, under the tailored classification procedure described above, some examinees may be classified as belonging either to State r or to State q with no way of distinguishing between the two. Two methods for dealing with this problem are proposed. Both methods involve first applying the modified classification procedure described above, and then applying an additional selection criterion only if the

examinee has been classified as belonging to two or more states that are indistinguishable with respect to the subset of items administered.

The first method proposed for dealing with the problem of indistinguishable states (such as States r and q above, if Item 5 were not administered) is appropriate when the primary purpose of the diagnostic procedure is to select a remediation program for the examinee. Under this method, the examinee is assigned to one or another of the possible states by selecting that state which indicates the least number of attributes mastered. In the example listed above, the examinee would be classified into State r. Note that this method assumes that the loss of providing remediation when remediation is not required is less than the loss of failing to remediate when remediation is required.

The second method proposed for dealing with the problem of indistinguishable states is appropriate when remediation is not the primary concern or when the losses associated with the two types of remediation errors are assumed to be equal. In this method, final classification decisions are made by comparing the prior probabilities associated with each of the possible states. In the example listed above, the examinee would be classified into State r or State q depending on which had the higher prior probability. The rationale for using prior probabilities to compare states derives from the result that, conditional on a previous classification to a cluster of indistinguishable states, the posterior probabilities of all states in that cluster are proportional to their prior probabilities. A proof of this result is given in Appendix A.

### **An Application to the Domain of Document Literacy**

The procedures outlined above have been applied to the document literacy data collected in the Survey of Young Adult Literacy, a nation-wide survey of literacy skills conducted by NAEP in 1985. This dataset includes 61 items classified as measuring document literacy, that is, the knowledge and skills needed to process information stored in non-prose formats such as tables, charts, or schedules (Kirsch and Jungeblut, 1986). These items were administered by trained interviewers: the examinee was handed a document, such as a page from a phone book or bus schedule, and was then asked to respond to one or two questions which required processing of at least some of the information stored in the document. The cognitive model assumed to be underlying performance in this domain was adapted from the work of Kirsch and Mosenthal (1990) who identified features of the items which were later shown to be highly correlated with the IRT difficulty parameters of the items (Sheehan and Mislevy, 1990).

The item feature variables identified by Kirsch and Mosenthal are listed in Table 2. These variables were originally measured on an ordinal scale. We have translated them into a set of 22 dichotomously scored attributes by coding the incidence matrix as indicated in Table 2. To illustrate this procedure, consider the coding listed for the Degree of Correspondence variable. This variable measures the degree to which the phrasing in the stem portion of the

item matches the phrasing in the document which the item refers to. It is scored on a 1 to 5 scale with lower values indicating more direct correspondence and thus, less difficulty; and higher values indicating less direct correspondence and thus, more difficulty. The first three ordered levels were translated into a set of three dichotomously scored attributes as follows: if an item is classified as requiring level 1 correspondence skills then an examinee would have to have mastered attribute C1 in order to correctly solve that item; if an item is classified as requiring level 2 skills then an examinee would have to have mastered attributes C1 and C2 in order to correctly solve that item; if an item is classified as requiring level 3 skills then an examinee would have to have mastered attributes C1, C2 and C3 in order to correctly solve that item. Levels 4 and 5 are translated analogously. Thus, the order relationships inherent in the ordinal levels of the original variables have been translated into order relationships among the attributes through the coding of the incidence matrix.

---

---

Insert Table 2 Here

---

---

Note that, under this coding scheme, it is impossible for an examinee to have mastered attribute C5 without also having mastered attributes C1 through C4. Similar restrictions apply to the other attributes. Thus, the attributes are now hierarchically ordered. This hierarchical ordering of the attributes is responsible for reducing the number of valid states from  $2^{22}$  to 7,776 or  $6 \times 6 \times 3 \times 3 \times 6 \times 4$ . The final number of valid states is much lower, however, since the item pool does not test all hierarchically-valid combinations of the attributes. That is, in the particular item pool developed for the NAEP literacy survey, items requiring medium to high mastery levels on some cognitive variables tended to also require medium to high mastery levels on other cognitive variables. Similarly, items requiring medium to low mastery levels on some cognitive variables tended to also require medium to low mastery levels on other cognitive variables. Since most combinations were not represented in the item pool (for example, Correspondence at Level 1 and Distractor at Level 5), the procedure for determining the subset of latent cognitive states to be considered found only 157 valid states.

The nationally representative adult literacy sample included approximately 3,600 scientifically selected examinees in the 21 to 25 age group. The subset of items presented to each examinee was determined through a BIB item sampling design in which the item pool was first divided into seven nonoverlapping blocks, and subsets consisting of three different blocks were subsequently arranged into seven distinct booklets such that each pair of blocks appeared together in exactly one booklet. The booklets were then spiralled into the population so that each booklet was administered to a random subsample of approximately 500 examinees. Because the original blocks differed in the number of document items they contained, the number of items in the resulting booklets also differed: from a low of 19 to a high of 41. These data were modeled using a two parameter logistic IRT model. Although item parameters were estimated using all of the available data, only those booklets which contained 30 or more items were included in the subset of data used to develop the diagnostic model. Booklets containing fewer than 30 items were excluded because  $\theta$  estimates based on



fewer than 30 items were considered to be too imprecise for use in classification. The final sample included three booklets, or three random subsamples containing a total of 1,509 examinees.

The projection of examinee response vectors into the two-dimensional rule space is presented in Figure 1. Examinees'  $\theta$  values are plotted along the x-axis, examinees'  $\zeta$  values are plotted along the y-axis. The plot shows a scatter of points in the  $\theta$  range from -3 to 3 and the  $\zeta$  range from -3 to 3. Figure 2 provides the projection of the 157 latent cognitive states into the rule space. As can be seen, there are very few states in the high  $\theta$  region. Thus, we should not expect to find high classification rates among high proficiency examinees. Figure 3 shows the prior probabilities assumed for each state. Prior probabilities were assumed to be proportional to the height of the bivariate normal density with mean (0,0) and covariance matrix equal to the identity. This prior was selected because (a) item parameters were estimated under the constraint of a standardized population distribution of  $\theta$ ; and (b) since  $\zeta$  is defined in standardized form, it is also expected to have a mean of zero and a standard deviation of one, whenever the IRT model fits.

---

---

Insert Figures 1, 2 and 3 Here

---

---

Using the procedure described previously (with an  $\alpha$ -level of .10), an admissibility region was determined for each examinee. A Bayes decision rule was then used to classify examinees into their "most possible" state. The classification results are summarized by classification outcome category in Table 3. The results show that 40% of the examinees were classified into a unique state, an additional 33% were classified into a set of two indistinguishable states, an additional 13% were classified into a set of three indistinguishable states, and so on. Overall, 90-percent of the examinees were classified into one or more of the 157 states. The fact that large numbers of examinees were not classified into a unique state indicates that the subset of items administered to each examinee did not test all of the relevant skills. This problem can be ameliorated in future document literacy assessments by specifying skill coverage as one of the characteristics to be considered in defining item subsets.

---

---

Insert Table 3 Here

---

---

Table 3 also lists the average number of items completed by an examinee in each classification outcome category. These values show that the probability of being classified into a unique state increases with the number of items completed. Note however that the 147 examinees who were not classified also completed a large number of items. This indicates that the classification failure was not due to insufficient data, but rather, to the fact that these examinees were responding in ways which were not consistent with the assumed cognitive

model. Thus, the cognitive model accounts for the document processing behaviors of only 90-percent of the population.

The number and percent of classified examinees is summarized by proficiency group and gender in Table 4. The low, medium and high proficiency groups were defined by dividing the original data set into thirds according to examinee's estimated  $\theta$  values. Thus, the 503 examinees with the lowest  $\theta$  values were classified into the low proficiency group, the 503 examinees with the highest  $\theta$  values were classified into the high proficiency group, and the remaining examinees were classified into the medium proficiency group. The table shows that the model works best for low proficiency examinees (95% classified) as opposed to medium or high proficiency examinees (88% classified). The breakdown by gender shows that females are more likely than males to be classified (93% as opposed to 87%).

---

---

Insert Table 4 Here

---

---

### Analysis of Attribute Mastery Probabilities

A vector of attribute mastery probabilities can be estimated for each classified examinee. For those examinees who were classified into a unique state (as was the case for 600 examinees in our sample) the probability of mastering any particular attribute will be either zero or one, depending on whether that attribute was included in the subset of attributes mastered defined for that state. (Note that we are ignoring the issue of classification error here. That issue is treated briefly at the end of this section.) When an examinee has been classified as belonging to a subset of two or more indistinguishable states, then the examinee's vector of attribute mastery probabilities can be determined by taking a weighted average of the attribute mastery probabilities defined for each state in the subset. Weights are selected to be proportional to the states' prior probabilities since, as was described previously, the posterior probability of each state in the subset is proportional to its prior probability. To illustrate this calculation, consider a cognitive model consisting of three attributes  $\{A_1, A_2, A_3\}$ , and an examinee who has been classified as belonging either to State  $r$  or to State  $q$ , where States  $r$  and  $q$  have the following subsets of attributes mastered: {State  $r$ :  $A_1$ }, and {State  $q$ :  $A_1, A_2$ }. The vector of attribute mastery probabilities for this examinee is calculated as follows:

$$p(A_1) = 1.0$$

$$p(A_2) = P(q)/[P(r) + P(q)]$$

$$p(A_3) = 0.0$$

where  $P(r)$  and  $P(q)$  represent prior probabilities for States  $r$  and  $q$ , respectively. Note that this procedure does not require us to select a unique "best" state for the examinee.

This method of calculating attribute mastery probabilities was applied to each of the 1,362 examinees who were classified in this study. The resulting attribute mastery probabilities were classified by proficiency group and gender and then analysed using a multivariate repeated measures analysis of variance, as described for instance in Myers (1979). A standard analysis of variance would not have been appropriate for these data because the hypothesis of multisample sphericity is violated. The results of this analysis are summarized in Table 5. (For reasons described below, the results given in Table 5 are based on 15 rather than 22 attributes.)

---

---

Insert Table 5 Here

---

---

The analysis of variance results reported in Table 5 provide evidence of three significant effects: proficiency group, attributes, and the attribute by proficiency group interaction. These results indicate that the attributes are differentially difficult and that examinees in different proficiency groups tend to have different attribute mastery profiles. The nonsignificance of the gender effects is interesting because it indicates that, for each attribute analysed, the average probability of mastery values calculated for males and females were very similar. Thus, the data provide no evidence of a gender difference in mastery of elementary document processing skills.

Table 6 presents the mean probability of mastery values estimated for each attribute. The different attribute mastery profiles obtained for low, medium and high proficiency examinees are clearly illustrated. The differential difficulty of the attributes is also shown. Note that, for each variable, the lowest classification level is mastered with a probability of 1.0 by examinees in all three proficiency groups. Thus, there is strong justification for excluding level 1 items from future document literacy assessments. Another thing to note is that attributes C3 and C4 have equal attribute mastery values in all three proficiency groups. This result is due to the fact that the item pool did not contain any items classified as level 3 on the correspondence variable. Thus, the probabilities listed for attribute C3 are no more than an artifact of the coding scheme developed for the incidence matrix. Because we have no valid information about mastery probabilities for attribute C3, and because we know for sure that all examinees have mastered attributes C1, D1, I1, O1, S1, and T1, these seven attributes were not included in the analysis of variance described previously.

---

---

Insert Table 6 Here

---

---

The last column in Table 6 provides the mean probability of mastery values estimated for the total sample of examinees. These values were obtained by taking an unweighted average of the mean values estimated in each of the three proficiency groups. Differences in these means were investigated using the multiple pairwise comparisons procedure described in Keselman, Keselman and Shaffer (1991). This procedure is appropriate because it uses estimates of variance for each comparison that are unbiased under violation of multisample sphericity. Using an overall  $\alpha$ -level of .05, four clusters of similarly difficult attributes were identified: {C5, D5}, {S3, D3, C2}, {D3, C2, O2}, {C2, O2, I4} and {I3, T2, I2, S2}. One thing to note about these clusters is that, except for I3 and I2, different levels of the same variable never appear together in the same cluster. Thus, for most variables, collapsing of levels is not indicated.

An alternative procedure for determining attribute mastery probabilities involves taking a weighted average of the attribute mastery designations defined for each state in the examinee's admissibility region. Although this alternative procedure was not used in this paper, we wish to note that it allows for an explicit treatment of classification error since weights may be defined to be proportional to states' posterior probabilities.

### **A Tree Representation of the Classification Results**

Often, diagnostic classification models are used to route examinees through computerized instructional systems. To assist in that purpose, this section presents a tree representation of the classification results obtained in this study.

The first step in devising a tree representation for a set of classification results involves selecting a single "best" state for each examinee who was classified into a subset of two or more states which were found to be indistinguishable with respect to the subset of items administered. As indicated earlier, this can be done by assigning examinees to states based on a loss function approach or by comparing states' prior probabilities. Because the primary purpose of the tree representation is to assist in routing examinees through computerized instructional systems, the loss function approach is the natural choice. This approach was applied to the document literacy classification results by assigning examinees to states such that the resulting classification indicated the least number of attributes mastered.

After all examinees have been assigned to their single "best" state, a subset of states which accounts for a large portion of the classified examinees must be determined. The subset of states selected for the document literacy tree representation consisted of all states with an observed frequency of seven or more examinees. This subset included 30 states and accounted for 92% of the classified examinees. The states included in this subset are listed in Table 7. The table also provides the attribute mastery designations for each state. As expected, states with high  $\theta$  values tend to have lots of mastered attributes and states with low  $\theta$  values tend to have fewer mastered attributes. The column of state frequencies shows

that this subset of states accounts for a total of 1,249 examinees, or 83% of the original sample.

---

---

Insert Table 7 Here

---

---

To develop a tree representation of the data given in Table 7, we start by plotting each state as a node and then draw arcs from one node to another, or from one state to another, to indicate transition relationships among the states. A transition from one state to another is said to be possible whenever the set of attributes associated with the first state is the largest available subset of the set of attributes associated with the second state. Thus, arcs connect lower states to higher states, where a higher state is defined as a state having at least one more attribute mastered. In some instances, of course, the next higher state will have two or more additional attributes mastered. The tree representation of the document processing classification results is given in Figure 4.

---

---

Insert Figure 4 Here

---

---

The node labels in Figure 4 identify the subset of attributes which would not be mastered by an examinee in the corresponding knowledge state. Thus, an examinee who is classified as having mastered all attributes except Correspondence Level 5 and Distractor Levels 4 and 5 would be assigned to the node labeled "C5,D4". The alternative remediation strategies available for this examinee are indicated by the two paths from node "C5,D4" to the state of perfect knowledge (represented by the blank node at the top of the figure). Path 1 progresses from "C5,D4" to "C5" and then to the blank node; Path 2 progresses from "C5,D4" to "D4", then to "D5" and then to the blank node. Path 1 corresponds to a remediation strategy in which the two distractor attributes are remediated first; Path 2 corresponds to a remediation strategy in which the correspondence attribute is remediated first. One way to choose between these two alternative remediation strategies is to consider the frequency values listed in Table 7. Path 1 has a frequency of 7 (7 examinees located at node "C5"); Path 2 has a frequency of 83 (59 examinees located at node "D4" and an additional 24 examinees located at node "D5"). Thus it is much more likely for an examinee to have mastered attribute "C5" before having mastered attributes "D4" and "D5" than the other way around. This suggests that a remediation strategy based on Path 2 has a higher probability of success than one based on Path 1.



## Discussion

This paper has shown that the Rule Space approach to diagnostic classification can be satisfactorily applied to data sets containing large amounts of missing data. With respect to the analysis of the NAEP document literacy data, there are three major findings to report:

- (1) For 40% of examinees, the Rule Space approach provided a precise diagnostic classification. That is, it indicated the particular subset of elementary document processing skills mastered by each examinee.
- (2) For an additional 33% of examinees, information about skill mastery was narrowed down to a set of two indistinguishable states. By comparing the attribute response vectors associated with each of these states, it would be possible to identify, for each examinee, the subset of skills known to be mastered, the subset skills known not to be mastered, and the subset of skills with mastery status still in question. A subsequent test could then be tailored to test only those skills which were still in question.
- (3) The data provide no evidence of a gender difference in mastery of elementary document processing skills.

In closing, we wish to note that two aspects of the document literacy application were somewhat atypical. First, all of the attributes were hierarchically ordered. Although the hierarchical ordering of attributes was responsible for a large reduction in the number of valid knowledge states, it was not necessary for application of the Rule Space approach. The only characteristics of attributes which are required for application of these procedures are: (1) they must be readily dichotomized and (2) they must be diagnostically relevant. Hierarchical ordering of the attributes will only come into play when the original variables are expressed on an ordinal or an interval scale.

The document literacy application was also atypical in that the problem of indistinguishable states was so pronounced. We wish to emphasize that the missing data would not have led to so many indistinguishable states if the cognitive characteristics of the items had been considered during the process of constructing item subsets.

## References

- Birenbaum, Menucha, Kelly, Anthony E., & Tatsuoka, Kikumi K. Toward a stable diagnostic representation of students' errors in algebra. (ETS Report No. RR-92-58-ONR). Princeton, NJ: Educational Testing Service.
- Dibello, L. V. & Baillee, R. J. (1991). Separating points in Rule Space: A contribution to diagnostic testing theory (Technical Report). Urbana, IL: University of Illinois, CERL.
- Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica*, 37, 359-374.
- Goodman, L. A. (1974). The analysis of qualitative variables when some of the variables are unobservable. Part I - A modified latent structure approach. *American Journal of Sociology*, 79, 1179-1259.
- Keselman, H. J., Keselman, J. C., & Shaffer, J. P. (1991). Multiple pairwise comparisons of repeated measures means under violation of multisample sphericity. *Psychological Bulletin*, 110, 162-170.
- Kirsch, I. S., & Jungeblut, A. (1986). *Literacy: Profiles of America's young adults* (Final Report No. 16-PL-01). Princeton, NJ: Educational Testing Service.
- Kirsch, I. S., & Mosenthal, P.B. (1990). Exploring document literacy: Variables underlying the performance of young adults. *Reading Research Quarterly*, 25, 5-29.
- Lazarsfeld, P. F. (1960). Latent structure analysis and test theory. In H. Gulliksen and S. Messick (Eds.), *Psychological Scaling, Theory and Applications*. New York: Wiley, 83-96.
- Macready, G. B. & Dayton, C. M. (1980). The nature and use of state mastery models. *Applied Psychological Measurement*, 4, 493-516.
- Mislevy, R. J., Beaton, A. E., Kaplan, B. & Sheehan, K. M. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement*, 29(2), 133-161.
- Myers, J.L. (1979). Fundamentals of Experimental Design (3<sup>rd</sup> ed.) Boston, MA: Allyn and Bacon.

- Scheiblechner, H. (1972). Das lernen und losen komplexer denkaufgaben. *Zeitschrift fur experimentalle und Angewandte Psychologie*, 19, 476-506.
- Sheehan, K. and Mislevy, R. J. (1990). Integrating cognitive and psychometric models to measure document literacy. *Journal of Educational Measurement*, 27(3), 255-272.
- Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, 20(4), 345-354.
- Tatsuoka, K. K. (1984). Caution indices based on item response theory. *Psychometrika*, 49, 95-110.
- Tatsuoka, K. K. (1985). A probabilistic model for diagnosing misconceptions in the pattern classification approach. *Journal of Educational Statistics*, 12(1), 55-73.
- Tatsuoka, K. K. & Tatsuoka, M. M. (1987). Bug distribution and statistical pattern classification. *Psychometrika*, 52, 193-206.
- Tatsuoka, K. K. (1991). *Boolean algebra applied to determination of the universal set of knowledge states* (Technical Report ONR-1). Educational Testing Service, Princeton, N.J.



Table 1

Sample  $\zeta$  Values  
 For Response Patterns with a Number-Correct Score of 3  
 from a Five-Item Test  
 With Rasch Item Difficulty Parameters of -2, -1, 0, 1, 2

Item Response Pattern <sup>a</sup>					$\zeta$
-2	-1	0	1	2	
1	1	1	0	0	-.85
1	1	0	1	0	.96
1	0	1	1	0	1.98
1	1	0	0	1	2.00
0	1	1	1	0	2.24
1	0	1	0	1	3.02
0	1	1	0	1	3.27
1	0	0	1	1	4.83
0	1	0	1	1	5.09
0	0	1	1	1	6.10

(a) All patterns yield  $\theta = .51$ .

Table 2

## The Document Literacy Variables &amp; Attributes

Variable Name / Level Description <sup>1</sup>	Attribute Name	Rows Coded 1 in the Inc. Matrix
<b>Degree of Correspondence between phrasing in the question or directive and in the document:</b>		
1) literal correspondence	C1	1
2) synonymous correspondence	C2	1,2
3) arrived at via low text-based inference	C3	1,2,3
4) arrived at via high text-based inference	C4	1,2,3,4
5) requires special prior knowledge	C5	1,2,3,4,5
<b>Type of Information processing required to identify and match features:</b>		
1) make a literal feature match	I1	6
2) make a low text-based inference	I2	6,7
3) make a high text-based inference	I3	6,7,8
4) make several conditional matches across nodes	I4	6,7,8,9
5) use special prior knowledge	I5	6,7,8,9,10
<b>No. of Organizing Categories (OCs) in the Directive:</b>		
1) 1 or less	O1	11
2) 2 or more	O2	11,12
<b>No. of Specifics in the Directive:</b>		
1) 2 or less	T1	13
2) 3 or more	T2	13,14
<b>Plausibility of Distractors:</b>		
1) no distractors	D1	15,
2) in same OC but do not share critical features	D2	15,16
3) in same OC and do share critical features	D3	15,16,17
4) appear in different OCs, at same level	D4	15,16,17,18
5) appear in different OCs, at different levels	D5	15,16,17,18,19
<b>No. of Specifics in the Document:</b>		
1) 50 or less	S1	20
2) between 51 and 100, inclusive	S2	20,21
3) greater than 100	S3	20,21,22

1. For complete level descriptions see Kirsch and Morenthal (1990).

**Table 3**

**The Initial Classification Results  
By Classification Outcome Category  
And Average Number of Items Completed**

No. of States	No. of Subjects	%	Avg. No. Items	Cum. No. Subjs	Cum. %
1	600	40	36.2	600	40
2	494	33	36.8	1094	73
3	203	13	33.2	1297	86
4	26	2	32.5	1323	88
>=5	39	3	25.7	1362	90
Not Class.	147	10	37.1	1509	100

No. of States = No. of states located at the selected point in the Rule Space.

**Table 4**

**The Number and Percent of Classified Examinees  
By Proficiency Group and Gender**

	<b>Total Subjects</b>	<b>No. Classified</b>	<b>Percent Classified</b>
<b>Proficiency Group</b>			
Low	503	476	95
Medium	503	443	88
High	503	443	88
<b>Gender Group</b>			
Female	845	787	93
Male	664	575	87
<b>All Subjects</b>	1509	1362	90

Table 5  
Analysis of Variance Results

Effect	Num. DF	Den. DF	F Value <sup>a</sup>	Pr>F
<b>Between Subjects</b>				
Proficiency	2	1356	655.44	.0001
Gender	1	1356	0.02	.8842
Prof X Gen	2	1356	0.64	.5295
<b>Within Subjects</b>				
Attributes	14	1343	1868.55	.0000
Att. X Prof.	28	2686	99.42	.0000
Att. X Gender	14	1343	0.75	.7280
Att X P X G	28	2686	1.00	.4711

(a) F values for within subject effects were calculated using Wilk's Lambda.

Table 6

Mean Attribute Mastery Probabilities

Att.	Proficiency			Total
	Low	Med	High	
C1	1.00	1.00	1.00	1.00
C2	0.68	0.87	1.00	0.85
C3	0.21	0.37	0.48	0.36
C4	0.21	0.37	0.48	0.36
C5	0.01	0.15	0.26	0.14
D1	1.00	1.00	1.00	1.00
D2	1.00	1.00	1.00	1.00
D3	0.70	0.85	1.00	0.85
D4	0.31	0.33	0.65	0.43
D5	0.13	0.16	0.24	0.17
I1	1.00	1.00	1.00	1.00
I2	0.94	1.00	1.00	0.98
I3	0.91	1.00	1.00	0.97
I4	0.68	0.98	1.00	0.88
I5	0.22	0.72	0.96	0.64
O1	1.00	1.00	1.00	1.00
O2	0.72	0.89	1.00	0.87
S1	1.00	1.00	1.00	1.00
S2	0.95	1.00	1.00	0.98
S3	0.56	0.90	1.00	0.82
T1	1.00	1.00	1.00	1.00
T2	0.91	1.00	1.00	0.97
All	0.64	0.75	0.82	0.74

Table 7

The Thirty Most Frequent States Ordered by  $\theta$ 

$\theta$	Freq.	Attributes Mastered
3.05	31	CCCCC IIIII OO TT DDDDD SSS
1.72	24	CCCCC IIIII OO TT DDDD- SSS
1.28	7	CCCC- IIIII OO TT DDDDD SSS
1.11	59	CCCCC IIIII OO TT DDD-- SSS
0.81	42	CC--- IIIII OO TT DDDDD SSS
.70	38	CCCC- IIIII OO TT DDD-- SSS
.62	102	CC--- IIIII OO TT DDDD- SSS
.39	296	CC--- IIIII OO TT DDD-- SSS
.33	18	CC--- IIII- OO TT DDDD- SSS
.29	8	CCCC- IIII- OO TT DDD-- SSS
.13	35	CC--- IIII- OO TT DDD-- SSS
-.23	64	CCCCC IIIII OO TT DD--- SSS
-.29	12	CC--- III-- OO TT DDD-- SSS
-.50	19	C---- IIIII OO TT DDDDD SSS
-.51	23	CCCC- IIIII O- TT DDDDD SSS
-.53	8	CC--- IIII- OO TT DD--- SSS
-.59	10	C---- IIIII OO TT DDDD- SSS
-.60	57	CCCC- IIIII O- TT DDD-- SSS
-.63	14	C---- IIIII OO TT DDD-- SSS
-.67	57	CC--- IIII- OO TT DDDD- SS-
-.67	42	CC--- IIIII O- TT DDDDD SSS
-.74	35	C---- IIII- OO TT DDDD- SSS
-.75	74	CC--- IIIII OO TT DDD-- SS-
-.78	23	C---- IIII- OO TT DDD-- SSS
-.92	38	C---- III-- OO TT DDD-- SSS
-1.06	13	CC--- IIII- OO TT DD--- SS-
-1.18	45	CC--- III-- OO TT DD--- SS-
-1.22	38	C---- III-- OO TT DD--- SSS
-1.61	9	CC--- I---- O- TT DD--- SS-
-2.03	8	CC--- I---- O- T- DD--- SS-
1249		

## Appendix A

**Proof that, Conditional on a Prior Classification to a Cluster of Indistinguishable States, the Posterior Probabilities of all States in the Cluster are Proportional to their Prior Probabilities.**

Let  $r$  and  $q$  be two states which are indistinguishable with respect to the subset of items administered. Let  $s$  represent the union of  $r$  and  $q$ . Let  $X$  represent an examinee's vector of observed item responses. (The number of elements in  $X$  will be less than the total number of items in the pool). Since  $r$  and  $q$  are indistinguishable we have

$$P(X|r) = P(X|q) = P(X|s)$$

The posterior probability of state  $r$ , conditional on a prior classification to state  $s$ , is calculated as

$$\begin{aligned} P(r|s, X) &= \frac{P(r \text{ and } s|X)}{P(s|X)} \\ &= \frac{P(r|X)}{P(s|X)} \\ &= \frac{P(X|r) P(r)}{P(X|s) P(s)} \\ &= \frac{P(r)}{P(s)} \end{aligned}$$

Similarly, the posterior probability of state  $q$ , conditional on a prior classification to state  $s$ , is

$$P(q|s, X) = \frac{P(q)}{P(s)}$$

Thus, conditional on a prior classification to a cluster of indistinguishable states, the posterior probability of any state in the cluster is proportional to its prior probability.



## FIGURE CAPTIONS

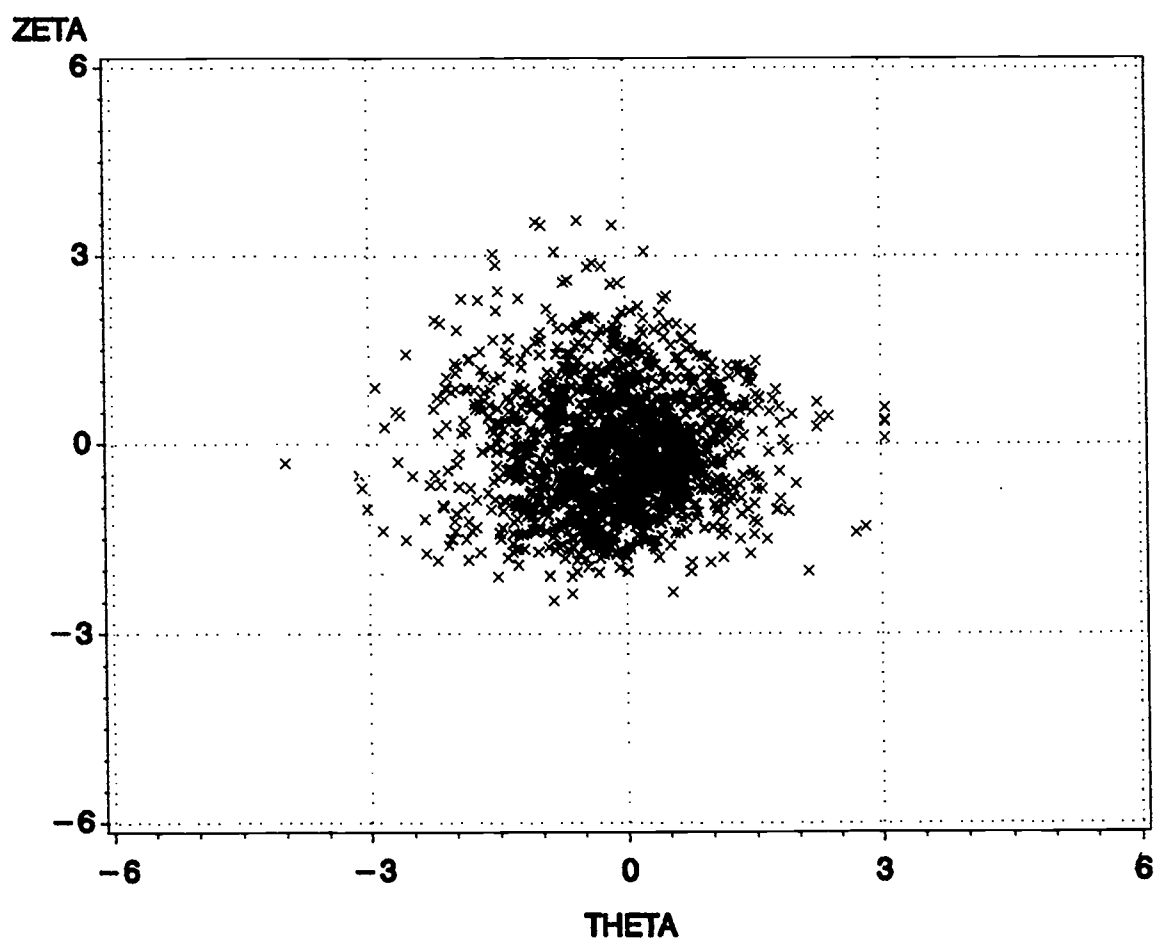
**Figure 1.** Projection of Examinee Response Data into the Rule Space.

**Figure 2.** Projection of the 157 states into the Rule Space.

**Figure 3.** Prior probabilities for the 157 states.

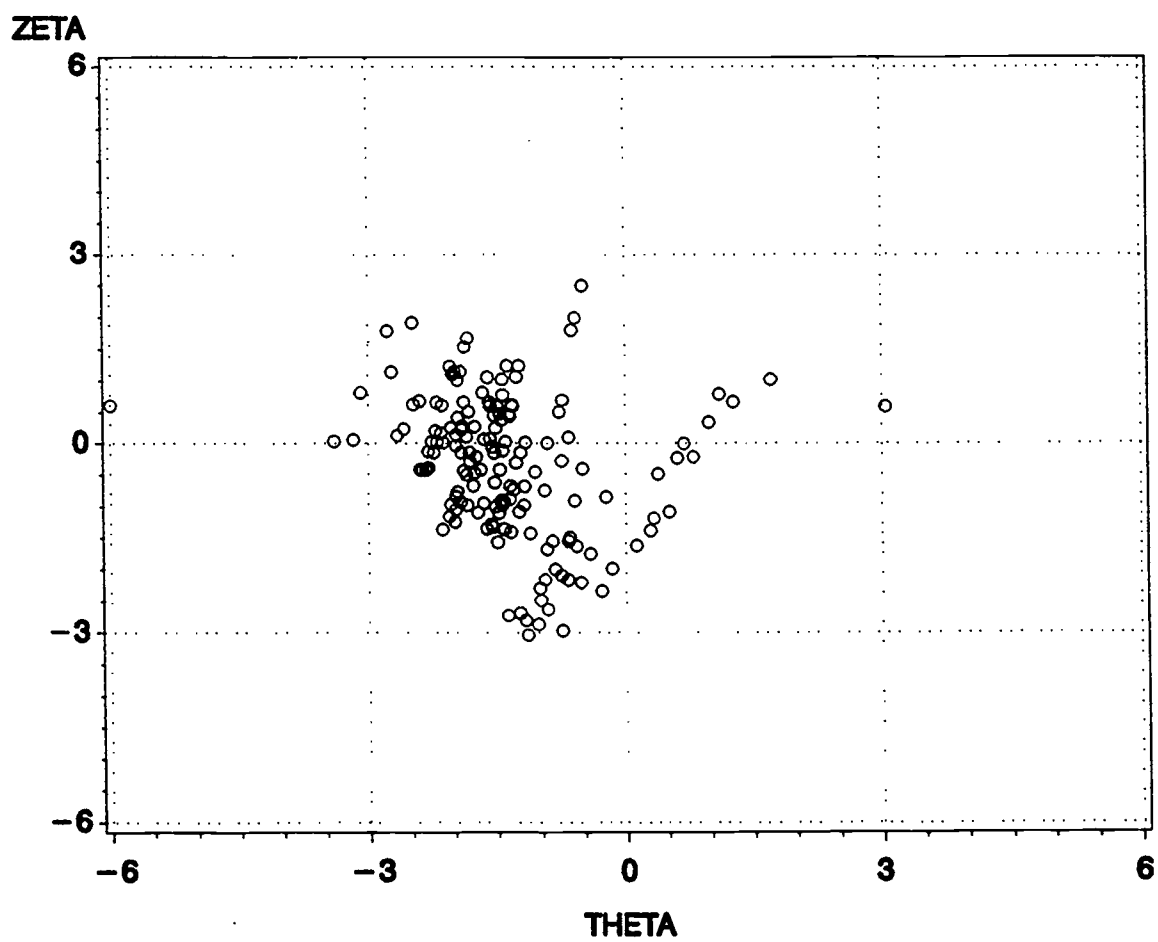
**Figure 4.** A Tree Representation of the classification results.

**Figure 1**  
**PROJECTION OF EXAMINEE RESPONSE DATA**  
**INTO THE RULE SPACE**



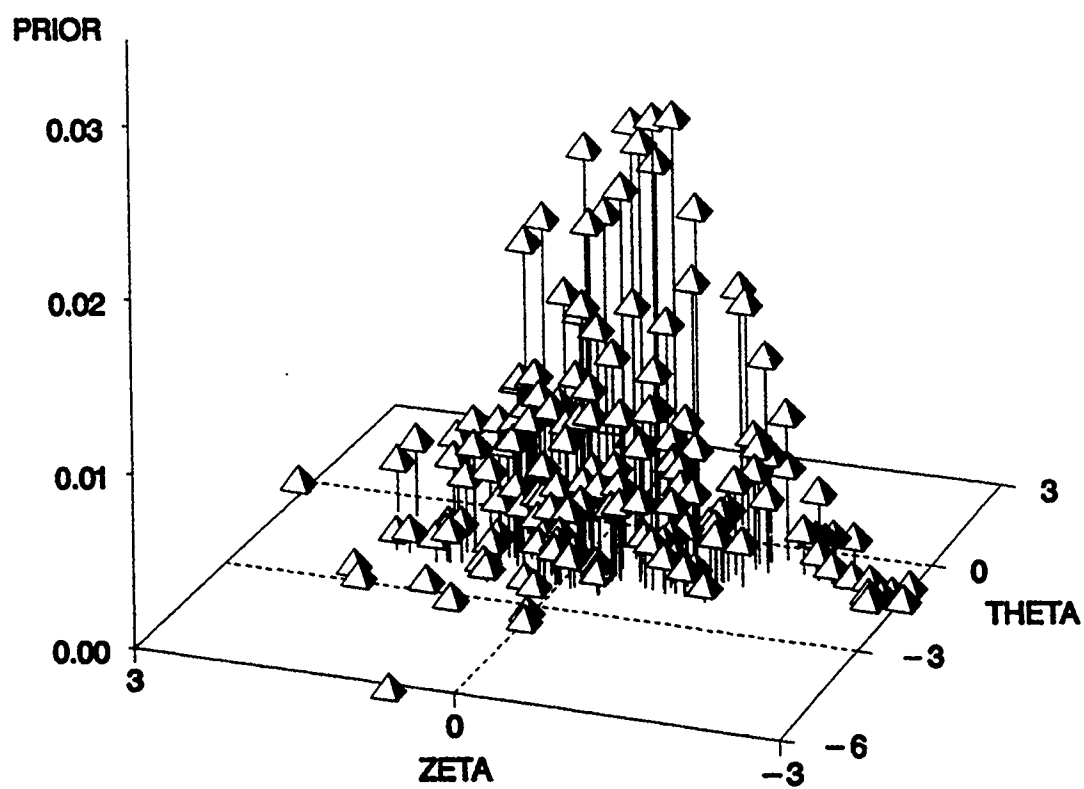
**Figure 2**

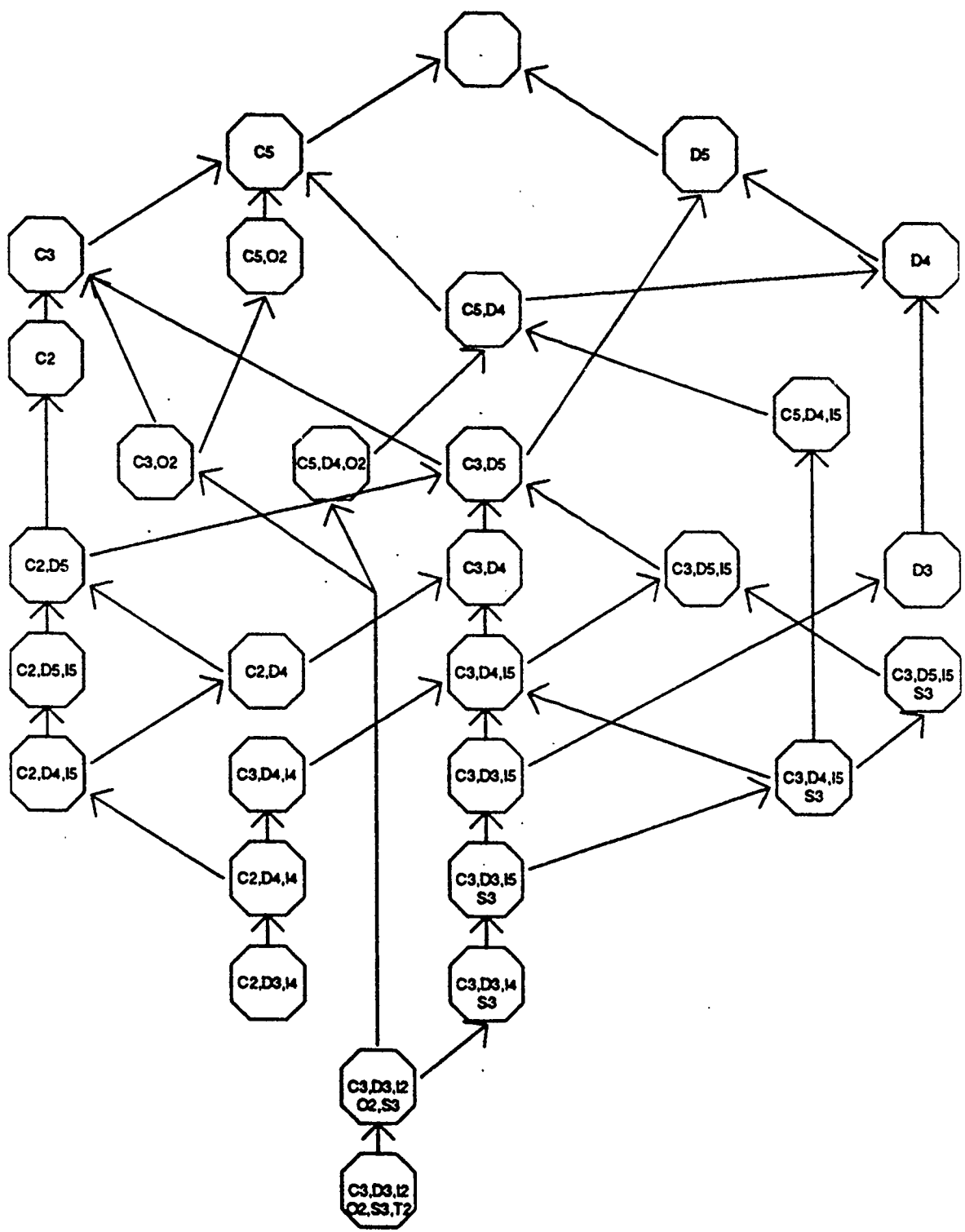
**PROJECTION OF THE 157 STATES  
INTO THE RULE SPACE**



**Figure 3**

**PRIOR PROBABILITIES  
FOR THE 157 STATES**





A Tree Representation of the Classification Results

Brophy 15 October 93

Distribution List

Dr Terry Ackerman  
Educational Psychology  
260C Education Bldg  
University of Illinois  
Champaign IL 61801

Dr Terry Allard  
Code 3422  
Office of Naval Research  
800 N Quincy St  
Arlington VA 22217-5660

Dr Nancy Allen  
Educational Testing Service  
Mail Stop 02-T  
Princeton NJ 08541

Dr Gregory Anrig  
Educational Testing Service  
Mail Stop 14-C  
Princeton NJ 08541

Dr Phipps Arabic  
Graduate School of Management  
Rutgers University  
92 New Street  
Newark NJ 07102-1895

Dr Isaac I Bejar  
Educational Testing Service  
Mail Stop 11-R  
Princeton NJ 08541

Dr William O Berry  
Director  
Life and Environmental Sciences  
AFOSR/NL N1  
Bldg 410  
Bolling AFB DC 20332-6448

Dr Thomas G Bever  
Department of Psychology  
University of Rochester  
River Station  
Rochester NY 14627

Dr Menucha Birenbaum  
Educational Testing Service  
Princeton NJ 08541

Dr Bruce Bloxom  
Defense Manpower Data Center  
99 Pacific St  
Suite 155A  
Monterey CA 93943-3231

Dr Gwyneth Boodoo  
Educational Testing Service  
Mail Stop 03-T  
Princeton NJ 08541

Dr Richard L Branch  
HQ USMEPCOM/MEPCT  
2500 Green Bay Road  
North Chicago IL 60064

Dr Robert Brennan  
American College Testing  
2201 North Dodge Street  
PO Box 168  
Iowa City IA 52243

Dr David V Budescu  
Department of Psychology  
University of Haifa  
Mount Carmel Haifa 31999  
ISRAEL

Dr Gregory Candell  
CTB/MacMillan/McGraw-Hill  
2500 Garden Road  
Monterey CA 93940

Dr Paul R Chatelier  
PERCEPTRONICS  
1911 North Ft Myer Drive  
Suite 1100  
Arlington VA 22209

Dr Susan Chipman  
Cognitive Science Program  
Office of Naval Research  
800 North Quincy Street  
Code 3422  
Arlington VA 22217-5660

Dr Raymond E Christal  
UES LAMP Science Advisor  
AL/HRMIL  
Brooks AFB TX 78235

Dr Norman Cliff  
Department of Psychology  
University of Southern California  
Los Angeles CA 90089-1061

Director  
Life Sciences  
Code 3420  
Office of Naval Research  
Arlington VA 22217-5660

Commanding Officer  
Naval Research Laboratory  
Code 4827  
Washington DC 20375-5000

Dr John M Cornwell  
Department of Psychology  
I/O Psychology Program  
Tulane University  
New Orleans LA 70118

Dr William Crano  
Department of Psychology  
Texas A&M University  
College Station TX 77843

Dr Linda Curran  
Defense Manpower Data Center  
Suite 400  
1600 Wilson Blvd  
Rosslyn VA 22209

Professor Clément Dassa  
Faculté des sciences de l'éducation  
Département d'études en éducation  
et d'administration de l'éducation  
CP 6128 succursale A  
Montréal Québec  
CANADA H3C 3J7

Dr Timothy Davey  
American College Testing  
2201 North Dodge Street  
PO Box 168  
Iowa City IA 52243

Dr Charles E Davis  
Educational Testing Service  
Mail Stop 16-T  
Princeton NJ 08541

Dr Ralph J DeAyala  
Meas Stat and Eval  
Benjamin Bldg Room 1230F  
University of Maryland  
College Park MD 20742

Dr Sharon Derry  
Florida State University  
Department of Psychology  
Tallahassee FL 32306

Hei-Ki Dong  
BELLCORE  
6 Corporate Place  
RM: PYA-1K207  
PO Box 1320  
Piscataway NJ 08855-1320

Dr Neil Dorans  
Educational Testing Service  
Mail Stop 07-E  
Princeton NJ 08541

Dr Fritz Drasgow  
University of Illinois  
Department of Psychology  
603 E Daniel Street  
Champaign IL 61820

Defense Tech Information Center  
Cameron Station Bldg 5  
Alexandria VA 22314  
(2 Copies)

Dr Richard Duran  
Graduate School of Education  
University of California  
Santa Barbara CA 93106

Dr Susan Embretson  
University of Kansas  
Psychology Department  
426 Fraser  
Lawrence KS 66045

Dr George Engelhard Jr  
Division of Educational Studies  
Emory University  
210 Fishburne Bldg  
Atlanta GA 30322

ERIC Facility-Acquisitions  
2440 Research Blvd  
Suite 550  
Rockville MD 20850-3238

Dr Marshall J Farr  
Farr-Sight Co  
2520 North Vernon Street  
Arlington VA 22207

Dr Leonard Feldt  
Lindquist Center for Measurement  
University of Iowa  
Iowa City IA 52242

Dr Richard L Ferguson  
American College Testing  
2201 North Dodge Street  
PO Box 168  
Iowa City IA 52243

Dr Gerhard Fischer  
Liebiggasse 5  
A 1010 Vienna  
AUSTRIA

Dr Myron Fischl  
US Army Headquarters  
DAPE-HR  
The Pentagon  
Washington DC 20310-0300

Mr Paul Foley  
Navy Personnel R&D Center  
San Diego CA 92152-6800

Chair  
Department of Computer Science  
George Mason University  
Fairfax VA 22030

Dr Robert D Gibbons  
University of Illinois at Chicago  
NPI 909A M/C 913  
912 South Wood Street  
Chicago IL 60612

Dr Janice Gifford  
University of Massachusetts  
School of Education  
Amherst MA 01003

Dr Robert Glaser  
Learning Res & Development Cntr  
University of Pittsburgh  
3939 O'Hara Street  
Pittsburgh PA 15260

Dr Susan R Goldman  
Peabody College  
Box 45  
Vanderbilt University  
Nashville TN 37203

Dr Timothy Goldsmith  
Department of Psychology  
University of New Mexico  
Albuquerque NM 87131

Dr Sherrie Gott  
AFHRL/MOMJ  
Brooks AFB TX 78235-5601

Dr Bert Green  
Johns Hopkins University  
Department of Psychology  
Charles & 34th Street  
Baltimore MD 21218

Professor Edward Haertel  
School of Education  
Stanford University  
Stanford CA 94305-3096

Dr Ronald K Hambleton  
University of Massachusetts  
Lab of Psychom & Eval Res  
Hills South Room 152  
Amherst MA 01003

Dr Delwyn Harnisch  
University of Illinois  
51 Gerty Drive  
Champaign IL 61820

Dr Patrick R Harrison  
Computer Science Department  
US Naval Academy  
Annapolis MD 21402-5002

Ms Rebecca Hetter  
Navy Personnel R&D Center  
Code 13  
San Diego CA 92152-6800

Dr Thomas M Hirsch  
American College Testing  
2201 North Dodge Street  
PO Box 168  
Iowa City IA 52243

Professor Paul W Holland  
Div of Educ Psych & Quant  
Methods Prog  
Graduate School of Education  
4511 Tolman Hall  
University of California-Berkeley  
Berkeley CA 94720

Professor Lutz F Hornke  
Institut für Psychologie  
RWTH Aachen  
Jaegerstrasse 17/19  
D-5100 Aachen  
WEST GERMANY

Ms Julia S Hough  
Cambridge University Press  
40 West 20th Street  
New York NY 10011

Dr William Howell  
Chief Scientist  
AFHRL/CA  
Brooks AFB TX 78235-5601

Dr Huynh Huynh  
College of Education  
University of South Carolina  
Columbia SC 29208

Dr Martin J Ippel  
Center for the Study of  
Education and Instruction  
Leiden University  
PO Box 9555  
2300 RB Leiden  
THE NETHERLANDS

Dr Robert Jannarone  
Elec and Computer Eng Dept  
University of South Carolina  
Columbia SC 29208

Dr Kumar Joag-dev  
University of Illinois  
Department of Statistics  
101 Illini Hall  
725 South Wright Street  
Champaign IL 61820

Professor Douglas H Jones  
Grad Sch of Management  
Rutgers The State University NJ  
Newark NJ 07102

Dr Brian Junker  
Carnegie-Mellon University  
Department of Statistics  
Pittsburgh PA 15213

Dr Marcel Just  
Carnegie-Mellon University  
Department of Psychology  
Schenley Park  
Pittsburgh PA 15213

Dr J L Kaiwi  
Code 442/JK  
Naval Ocean Systems Center  
San Diego CA 92152-5000

Dr Michael Kaplan  
Office of Basic Research  
US Army Research Institute  
5001 Eisenhower Avenue  
Alexandria VA 22333-5600

Dr Jeremy Kilpatrick  
Dept of Mathematics Education  
105 Aderhold Hall  
University of Georgia  
Athens GA 30602

Ms Hae-Rim Kim  
University of Illinois  
Department of Statistics  
101 Illini Hall  
725 South Wright Street  
Champaign IL 61820

Dr. Jwa-keun Kim  
Department of Psychology  
Middle Tennessee State University  
Murfreesboro TN 37132

Dr Sung-Hoon Kim  
KEDI  
92-6 Umyeon-Dong  
Seocho-Gu  
Seoul  
SOUTH KOREA

Dr G Gage Kingsbury  
Portland Public Schools  
Res & Eval Department  
501 North Dixon Street  
PO Box 3107  
Portland OR 97209-3107

Dr William Koch  
Box 7246  
Meas & Eval Center  
University of Texas-Austin  
Austin TX 78703

Dr James Kraatz  
Computer-based Education  
Research Laboratory  
University of Illinois  
Urbana IL 61801

Dr Patrick Kyllonen  
AFHRL/MOEL  
Brooks AFB TX 78235

Ms Carolyn Laney  
1515 Spencerville Rod  
Spencerville MD 20868

Richard Lanterman  
Commandant (G-PWP)  
US Coast Guard  
2100 Second Street SW  
Washington DC 20593-0001

Dr Michael Levine  
Educational Psychology  
210 Education Building  
1310 South Sixth Street  
Univ of IL at Urbana-Champaign  
Champaign IL 61820-6990



Dr Charles Lewis  
Educational Testing Service  
Mail Stop 03-T  
Princeton NJ 08541-0001

Mr Hsin-hung Li  
University of Illinois  
Department of Statistics  
101 Illini Hall  
725 South Wright Street  
Champaign IL 61820

Library  
Naval Training Systems Center  
12350 Research Parkway  
Orlando FL 32826-3224

Dr Marcia C Linn  
Graduate School of Education  
EMST  
Tolman Hall  
University of California  
Berkeley CA 94720

Dr Robert L Linn  
Campus Box 249  
University of Colorado  
Boulder CO 80309-0249

Logicon Inc (Attn: Library)  
Tactical & Training Systems Div  
PO Box 85158  
San Diego CA 92138-5158

Dr Richard Luecht  
American College Testing  
2201 North Dodge Street  
PO Box 168  
Iowa City IA 52243

Dr George B. Macready  
Dept of Meas Stat & Eval  
College of Education  
University of Maryland  
College Park MD 20742

Dr Evans Mandes  
George Mason University  
4400 University Drive  
Fairfax VA 22030

Dr Paul Mayberry  
Center for Naval Analysis  
4401 Ford Avenue  
PO Box 16268  
Alexandria VA 22302-0268

Dr James R McBride  
HumRRO  
6430 Elmhurst Drive  
San Diego CA 92120

Mr Christopher McCusker  
University of Illinois  
Department of Psychology  
603 E Daniel Street  
Champaign IL 61820

Dr Robert McKinley  
Educational Testing Service  
Princeton NJ 08541

Dr Joseph McLachlan  
Navy Pers Res & Dev Cntr  
Code 14  
San Diego CA 92152-6800

Alan Mead  
c/o Dr Michael Levine  
Educational Psychology  
210 Education Bldg  
University of Illinois  
Champaign IL 61801

Dr Timothy Miller  
American College Testing  
2201 North Dodge Street  
PO Box 168  
Iowa City IA 52243

Dr Robert Mislevy  
Educational Testing Service  
Mail Stop 03-T  
Princeton NJ 08541

Dr Ivo Molenaar  
Faculteit Sociale Wetenschappen  
Rijksuniversiteit Groningen  
Grote Kruisstraat 2/1  
9712 TS Groningen  
The NETHERLANDS

Dr Eiji Muraki  
Educational Testing Service  
Mail Stop 02-T  
Princeton NJ 08541

Dr Ratna Nandakumar  
Educational Studies  
Willard Hall Room 213E  
University of Delaware  
Newark DE 19716

Acad Prog & Research Branch  
Naval Tech Training Command  
Code N-62  
NAS Memphis (75)  
Millington TN 30854

Dr W Alan Nicewander  
American College Testing  
2201 North Dodge Street  
PO Box 168  
Iowa City IA 52243

Head  
Personnel Systems Department  
NPRDC (Code 12)  
San Diego CA 92152-6800

Director  
Training Systems Department  
NPRDC (Code 14)  
San Diego CA 92152-6800

Library NPRDC  
Code 041  
San Diego CA 92152-6800

Librarian  
Naval Cntr for Applied Research  
in Artificial Intelligence  
Naval Research Laboratory  
Code 5510  
Washington DC 20375-5000

Office of Naval Research  
Code 3422  
800 N Quincy Street  
Arlington VA 22217-5660  
(6 Copies)

ONR Resident Representative  
New York City  
33 Third Avenue - Lower Level  
New York NY 10003-9998

Special Asst for Res Management  
Chief of Naval Personnel  
(PERS-01JT)  
Department of the Navy  
Washington DC 20350-2000

Dr Judith Orasanu  
NASA Ames Research Center  
Mail Stop 239-1  
Moffett Field CA 94035

Dr Peter J Pashley  
Law School Admission Services  
PO Box 40  
Newtown PA 18840-0040

Wayne M Patience  
American Council on Education  
GED Testing Service Suite 20  
One Dupont Circle NW  
Washington DC 20036

Dept of Administrative Sciences  
Code 54  
Naval Postgraduate School  
Monterey CA 93943-5026

Dr Peter Pirolli  
School of Education  
University of California  
Berkeley CA 94720

Dr Mark D Reckase  
American College Testing  
2201 North Dodge Street  
PO Box 168  
Iowa City IA 52243

Mr Steve Reise  
Department of Psychology  
University of California  
Riverside CA 92521

Mr Louis Roussos  
University of Illinois  
Department of Statistics  
101 Illini Hall  
725 South Wright Street  
Champaign IL 61820

Dr Donald Rubin  
Statistics Department  
Science Center Room 608  
1 Oxford Street  
Harvard University  
Cambridge MA 02138

Dr Fumiko Samejima  
Department of Psychology  
University of Tennessee  
310B Austin Peay Bldg  
Knoxville TN 37966-0900

Dr Mary Schratz  
4100 Parkside  
Carlsbad CA 92008

Mr Robert Semmes  
N218 Elliott Hall  
Department of Psychology  
University of Minnesota  
Minneapolis MN 55455-0344

Dr Valerie L Shalin  
Dept of Industrial Engineering  
State University of New York  
342 Lawrence D Bell Hall  
Buffalo NY 14260

Mr Richard J Shavelson  
Graduate School of Education  
University of California  
Santa Barbara CA 93106

Kathleen Sheehan  
Educational Testing Service  
Mail Stop 03-T  
Princeton NJ 08541

Dr Kazuo Shigemasu  
7-9-24 Kugenuma-Kaigan  
Fujisawa 251  
JAPAN

Dr Randall Shumaker  
Naval Research Laboratory  
Code 5500  
4555 Overlook Avenue SW  
Washington DC 20375-5000

Dr Judy Spray  
American College Testing  
2201 North Dodge Street  
PO Box 168  
Iowa City IA 52243

Dr Martha Stocking  
Educational Testing Service  
Mail Stop 03-T  
Princeton NJ 08541

Dr William Stout  
University of Illinois  
Department of Statistics  
101 Illini Hall  
725 South Wright St  
Champaign IL 61820

Dr Kikumi Tatsuoka  
Educational Testing Service  
Mail Stop 03-T  
Princeton NJ 08541

Dr David Thissen  
Psychometric Laboratory  
CB# 3270 Davie Hall  
University of North Carolina  
Chapel Hill NC 27599-3270

Mr Thomas J Thomas  
Federal Express Corporation  
Human Resource Development  
3035 Director Row Suite 501  
Memphis TN 38131

Mr Gary Thomasson  
University of Illinois  
Educational Psychology  
Champaign IL 61820

Dr Howard Wainer  
Educational Testing Service  
15-T Rosedale Road  
Princeton NJ 08541

Elizabeth Wald  
Office of Naval Technology  
Code 227  
800 North Quincy Street  
Arlington VA 22217-5000

Dr Michael T Waller  
Univ of Wisconsin-Milwaukee  
Educ Psychology Department  
Box 413  
Milwaukee WI 53201

Dr Ming-Mei Wang  
Educational Testing Service  
Mail Stop 03-T  
Princeton NJ 08541

Dr Thomas A Warm  
FAA Academy  
PO Box 25082  
Oklahoma City OK 73125

Dr David J Weiss  
N660 Elliott Hall  
University of Minnesota  
75 E River Road  
Minneapolis MN 55455-0344

Dr Douglas Wetzel  
Code 15  
Navy Personnel R&D Center  
San Diego CA 92152-6800

German Military Representative  
Personalstammamt  
Koelner Str 262  
D-5000 Koeln 90  
WEST GERMANY

Dr David Wiley  
Sch of Educ and Social Policy  
Northwestern University  
Evanston IL 60208

Dr Bruce Williams  
Dept of Educational Psychology  
University of Illinois  
Urbana IL 61801

Dr Mark Wilson  
School of Education  
University of California  
Berkeley CA 94720

Dr Eugene Winograd  
Department of Psychology  
Emory University  
Atlanta GA 30322

Dr Martin F Wiskoff  
PERSEREC  
99 Pacific Street  
Suite 4556  
Monterey CA 93940

Mr John H Wolfe  
Navy Personnel R&D Center  
San Diego CA 92152-6800

Dr Kentaro Yamamoto  
Educational Testing Service  
Mail Stop 03-T  
Princeton NJ 08541

Duanli Yan  
Educational Testing Service  
Mail Stop 03-T  
Princeton NJ 08541

Dr Wendy Yen  
CTB/McGraw Hill  
Del Monte Research Park  
Monterey CA 93940

Dr Joseph L Young  
National Science Foundation  
Room 320  
1800 G Street NW  
Washington DC 20550